Conversion of graded to binary response in an activator-repressor system

Rajesh Karmakar*

January 19, 2010

Department of Physics

A. K. P. C. Mahavidyalaya

Subhasnagar, Bengai, Hooghly-712 611, India.

Abstract

Appropriate regulation of gene expression is essential to ensure that protein synthesis occurs in a selective manner. The control of transcription is the most dominant type of regulation mediated by a complex of molecules such as transcription factors. In general, regulatory molecules are of two types: activator and repressor. Activators promote the initiation of transcription whereas repressors inhibit transcription. In many cases, they regulate the gene transcription on binding the promoter mutually exclusively and the observed gene expression response is either graded or binary. In experiments, the gene expression response is quantified by the amount of proteins produced on varying the concentration of an external inducer molecules in the cell. In this paper, we study a gene regulatory network where activators and repressors both bind the same promoter mutually exclusively. The network is modeled by assuming that the gene can be in three possible states: repressed, unregulated and active. An exact analytical expression for the steady-state probability distribution of protein levels is then derived. The exact result helps to explain the experimental observations that in the presence of activator molecules the response is graded at all inducer levels whereas in the presence of both activator and repressor molecules, the response is graded at low and high inducer levels and binary at an intermediate inducer level.

PACS number(s): 87.10.Mn

I. INTRODUCTION

Gene expression, a fundamental cellular process whereby mRNAs and proteins are synthesized, is inherently stochastic in nature. There is a large number of theoretical and experimental studies which confirm the stochastic nature of gene expression [1]. The stochasticity or noise in gene expression is due to the small number of molecules involved in the associated cellular processes. For example, the DNA molecule which gives an organism its unique genetic identity is present in one or two copies per cell. The small number of molecules taking part in the biochemical events of

*Electronic address: rkarmakar2001@yahoo.com

gene expression is responsible for the probabilistic occurrence of the events. The stochastic nature of the biochemical events introduces fluctuations around the mean mRNA and protein levels. The fluctuations constitute noise and cause identical copies of a gene to express at different levels. The total noise in the gene expression level has two components: intrinsic and extrinsic. The origin of intrinsic noise lies in the probabilistic nature of the biochemical events of gene expression. The sources of extrinsic noise is in the fluctuations in cellular components such as RNAPs, ribosomes and regulatory molecules. The noise in gene expression may give rise to heterogeneity in a cell population. Cell-to-cell variability is generally attributed to genetic differences though the environment and history are also contributing factors. Recent experiments [2, 3] provide evidence that stochasticity in gene expression can contribute substantially to population heterogeneity and consequent variability in the cellular phenotype. A population of cells with identical genetic sequences as well as history and subjected to the same constant environment can develop heterogeneities due to the random nature of gene expression. Cellular heterogeneity has been observed in a variety of cell types ranging from bacteria [5] to complex mammalian cells [6]. Several experiments combined with theoretical studies provide important new insight on the stochastic aspects of gene expression [2, 3, 5, 7, 8, 9].

Gene expression and its regulation are of fundamental importance in living organisms. There are many steps in gene expression pathway from DNA to proteins and different types of regulatory molecules are involved in different steps. In general, transcriptional regulation is one of the most dominant types of regulation. The activator and repressor molecules are actively involved in the regulation of gene transcription both in prokaryotes and eukaryotes. Transcriptional repressors such as lac and tryptophan repressors are well known for prokaryotic systems. Repressor molecules inhibit the gene transcription by binding to the appropriate region of the DNA. Eukaryotic systems are much more complex and have compact chromatin structures. For the initiation of transcription, remodelling of the chromatin structure is essential so that the transcription factors and the RNA polymerase have access to the appropriate binding regions. Thus, gene activation in eukaryotic system means the relief of repression by the nucleosomal structure of the chromatin. After remodelling of chromatin structure, activator protein binds the DNA and activate gene expression. Activator protein concentrations can be varied by varying the inducer molecules such as galactose [3].

Experiments reveal that in an individual cell the gene expression response, the amount of proteins synthesized, can be of two types: graded and binary. In graded response protein level varies continuously with varying concentration of external inducer molecules. In binary response, protein levels can have two possible values: low or high. This is also known as the all-or-none phenomenon in gene expression. The binary response at the single cell level gives rise to a bimodal distribution in protein levels at the population level. There are experimental evidences of binary responses in gene expression with different possible origins [3, 10, 11, 12, 13, 14]. Becskei et al. [10] have demonstrated that positive feedback with cooperativity can generate binary response in a synthetic eukaryotic gene circuit but without the positive feedback the response is graded. The presence of positive feedback loop with cooperativity gives rise to bistability and the bistability along with stochasticity produces binary response in protein levels. Recently, Tan et al. [11] have established that bistability may also arises from the interplay between a non-cooperative positive feedback loop and circuit-induced growth retardation. Blake et al. [3] and Karmakar and Bose [4] have shown that fluctuations in the levels of transcription factor can give rise to binary responses in the target gene expression in an eukaryotic system. Rossi et al. [12] and Biggar and Carbtree [13] have further shown that, in certain instances, competition between activator and repressor molecules to occupy the promoter region can generate a binary response in gene expression. If the activator or repressor molecules act independently, a graded response is obtained. The difference

in the cellular fates in binary response may be ascribed to heterogeneity in the distribution of the stimulus/inducer molecules in the cell population, different histories, i.e., initial states in the case of bistability, intracellular noise giving rise to fluctuations in key parameter values etc.

To explain the experimentally observed binary responses in gene expression, different modeling approaches have already been proposed and analyzed using simulation and analytical techniques. Binary response in an autocatalytic induction circuit is very common and easily understood from different theoretical studies [15, 16]. Noise can have important role in the generation of binary responses in gene expression. It may be of purely stochastic origin [17, 18, 19]. Kepler and Elston [17] have demonstrated through specific examples that only stochasticity in gene expression can give rise to binary response, i.e., a bimodal distribution in the protein levels. Pirone and Elston [18] show that the slow promoter transition in gene states is responsible for binary responses whereas fast transitions produce graded responses. Karmakar and Bose [19] defined the slow and fast transitions between the active and inactive states of the gene more precisely and established the conditions of origin of graded and binary responses in gene expression. They derived the distribution of protein levels assuming the random transitions between the gene states with protein synthesis and degradation occurring deterministically. Later, exact analytical distributions for mRNAs and proteins have been derived considering all the major steps of gene expression i.e., transcription, translation and degradation, to be stochastic [20, 21]. In this paper, we propose a simple model of stochastic gene transcription regulated by activators and repressors and show using exact analytical calculations that bimodal distribution in protein levels appears naturally when activators and repressors compete for the binding site mutually exclusively to regulate the gene transcription. On the other hand, a graded response is observed when only activator molecules regulate the gene transcription.

II. STOCHASTIC MODEL AND EXACT SOLUTION

Transcriptional regulation by activator and repressor molecules on binding the same promoter is an important regulatory mechanism of gene expression in living organisms. The activator (repressor) molecules activate (inhibit) the transcription by binding the appropriate site on the promoter. Here we consider a gene regulatory network where activators and repressors both regulate the gene transcription mutually exclusively [12]. This can happen in different ways and one such way may be the overlapping binding sites on the promoter (Fig. 1). Therefore, the activator and repressor molecules cannot bind the promoter simultaneously, rather they compete for their binding sites to regulate gene transcription. This mechanism of transcriptional regulation is represented by a simple reaction scheme (Fig. 2) where a gene can be in three possible states: G_1, G_2 and G_3. G_2 is the unregulated state and G_1 (G_3) is the repressed (activated) state of the gene. The unregulated state of the gene which is achieved when both the sites are empty. Activator (repressor) molecules, on binding its specific site, help in transition from the unregulated state G_2 to the active (repressed) state G_3 (G_1) of the gene. There are random transitions taking place between the three states of the gene. Activator and repressor molecules compete for the state G_2 to take control of the network. If activator molecule wins, the gene turns into active state and protein synthesis occurs with rate constant J_p. Protein production does not take place from the unregulated (G_2) and repressed (G_1) states of the gene. Degradation of proteins occur with rate constant k_p and this event is independent of the states of the gene. Here transcription and translation are combined together into a single step as done in earlier studies [17, 19]. The stochastic transition from G_2 to G_3 occurs with rate constant k_a and that from G_2 to G_1 with k_2 (Fig. 2). The rate constants k_a and k_2 are the functions of activator and repressor molecules respectively. Thus, in absence

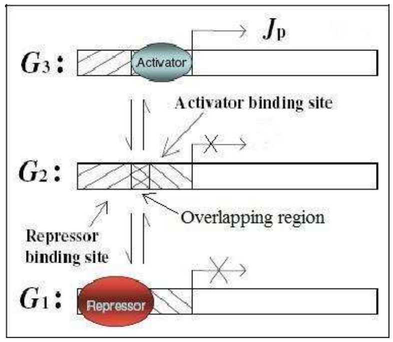

FIG. 1. Schematic diagram of transcriptional regulation by activator and repressor molecules where both the molecules compete for their respective binding site.

$$\mathbf{G_1} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathbf{G_2} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathbf{G_3} \xrightarrow{Jp} \mathbf{P} \xrightarrow{kp} \phi$$

FIG. 2. Reaction scheme with the three states of the gene: repressed (G_1), unregulated (G_2) and activated (G_3). From the activated state G_3 proteins are synthesized with rate constant J_p.

of repressor (activator) molecules the transition from G_2 to G_1 (G_3) is not possible at all. The assumption that there can be three possible states of the gene provides the basis for the minimal model of the activator-repressor system.

Let $p_i(n,t)$ $(i = 1,2,3)$ be the probability that at time t, the gene is in the G_i state with n number of protein molecules in the system. The Master equations for the biochemical reactions corresponding to the Fig. 2 are given by

$$\frac{\partial p_1(n,t)}{\partial t} = k_2\, p_2(n,t) - k_1\, p_1(n,t) + k_p[(n+1)p_1(n+1,t) - n\, p_1(n,t)] \tag{1}$$

$$\frac{\partial p_2(n,t)}{\partial t} = k_1\, p_1(n,t) + k_d\, p_3(n,t) - k_2\, p_2(n,t) - k_a\, p_2(n,t) + J_0[p_2(n-1,t) - p_2(n,t)]$$

$$+ k_p[(n+1)p_2(n+1,t) - n\, p_2(n,t)] \tag{2}$$

$$\frac{\partial p_3(n,t)}{\partial t} = k_a\, p_2(n,t) - k_d\, p_3(n,t) + J_p[p_3(n-1,t) - p_3(n,t)] + k_p[(n+1)p_3(n+1,t) - n\, p_3(n,t)] \tag{3}$$

Now the standard approach of the theory of stochastic processes will be used to determine the steady-state probability density function for protein levels [22]. The generating functions are defined as

$$F_1(z,\, t) = \sum_n z^n\, p_1(n,t),\;\; F_2(z,\, t) = \sum_n z^n\, p_2(n,t),\;\;\; F_3(z,\, t) = \textstyle\sum_n z^n\, p_3(n,t) \text{ and } F(z,t) = \displaystyle\sum_n z^n\, p(n,t) \tag{4}$$

where

$$\begin{aligned} F(z,t) &= F_1(z,t) + F_2(z,t) + F_3(z,t) \\ p(n,t) &= p_1(n,t) + p_2(n,t) + p_3(n,t) \end{aligned} \tag{5}$$

where $F(z,t)$ and $p(n,t)$ are the total generating function and total probability density function respectively.

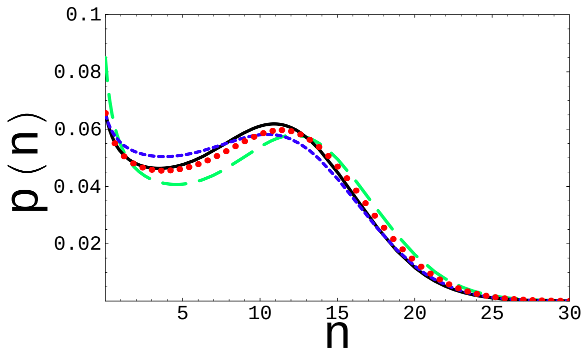

FIG. 3. Plot of $p(n)$ versus n for the activator-repressor system for $b_1 = 16$ and four different sets of parameter values: long dashed curve: $s_1 = 1$, $s_2 = 6$, $s_a = 10$, $s_d = 1$, for solid curve: $s_1 = 1$, $s_2 = 4$, $s_a = 13$, $s_d = 2$, for short dashed curve: $s_1 = 2$, $s_2 = 6$, $s_a = 5$, $s_d = 1$ and for dotted curve: $s_1 = 1.25$, $s_2 = 6$, $s_a = 10$, $s_d = 1.25$.

In terms of the generating functions (4), Eqs. (1), (2) and (3) can be written as

$$\frac{\partial F_1(z,t)}{\partial t} = k_2 F_2(z,t) - k_1 F_1(z,t) + k_p(1-z)\frac{\partial F_1(z,t)}{\partial z} \tag{6}$$

$$\frac{\partial F_2(z,t)}{\partial t} = k_1 F_1(z,t) + k_d F_3(z,t) - k_2 F_2(z,t) - k_a F_2(z,t) + k_p(1-z)\frac{\partial F_2(z,t)}{\partial z} \tag{7}$$

$$\frac{\partial F_3(z,t)}{\partial t} = k_a F_2(z,t) - k_d F_3(z,t) + J_p(z-1)F_3(z,t) + k_p(1-z)\frac{\partial F_3(z,t)}{\partial z} \tag{8}$$

In the steady state ($\frac{\partial F_i}{\partial t} = 0$, $i = 1, 2, 3$), addition of Eqs. (6), (7) and (8) results

$$J_p F_3(z) = k_p \frac{\partial F(z)}{\partial z} \tag{9}$$

With the help of the Eqs. (5), (6), (8) and (9), $F_1(z)$ and $F_2(z)$ can be expressed in terms of $F(z)$. Then, in terms of the generating function $F(z)$, the Eqs. (6), (7) and (8) can be written as

$$(z-1)^2 F^{'''}(z) + \{a_1(z-1) - b_1(z-1)^2\}F^{''}(z) + \{a_2 - b_1 b_2(z-1)\}F^{'}(z) - b_1 a_3 F(z) = 0 \tag{10}$$

where $a_1 = (1 + s_1 + s_2 + s_a + s_d)$, $b_1 = J_p/k_p$, $a_2 = s_1 s_a + s_1 s_d + s_2 s_d$, $b_2 = a_1 - s_d$, $a_3 = s_1 s_a$, $s_1 = k_1/k_p$, $s_2 = k_2/k_p$, $s_a = k_a/k_p$ and $s_d = k_d/k_p$.
The solution of the Eq. (10) is a generalized hypergeometric function and is given by

$$F(z) = C\,{}_pF_q[g_1 - g_2;\ g_1 + g_2;\ h_1 - h_2;\ h_1 + h_2;\ b_1(z-1)] \tag{11}$$

where $g_1 = -\frac{1}{2} + \frac{b_2}{2}$, $g_2 = \frac{1}{2}\sqrt{(b_2-1)^2 - 4a_3}$, $h_1 = -\frac{1}{2} + \frac{a_1}{2}$, $h_2 = \frac{1}{2}\sqrt{(a_1-1)^2 - 4a_2}$, C is the normalization constant and ${}_pF_q(a,b,c,d)$ is the generalized hypergeometric function (GHF). The normalization constant can be determined easily from the condition $F(1) = 1$.

Differentiating Eq. (11) n times w.r.t. z at $z = 0$, one can easily obtain the expression for the steady-state probability density function $p(n)$ as

$$p(n) = C\,\frac{b_1^n\,\Gamma(g_1+n)\,\Gamma(g_2+n)\,\Gamma(h_1)\,\Gamma(h_2)}{n!\,\Gamma(h_1+n)\,\Gamma(h_2+n)\,\Gamma(g_1)\,\Gamma(g_2)}\,{}_pF_q(g_1+n;\ g_2+n;\ h_1+n;\ h_2+n;\ -b_1) \tag{12}$$

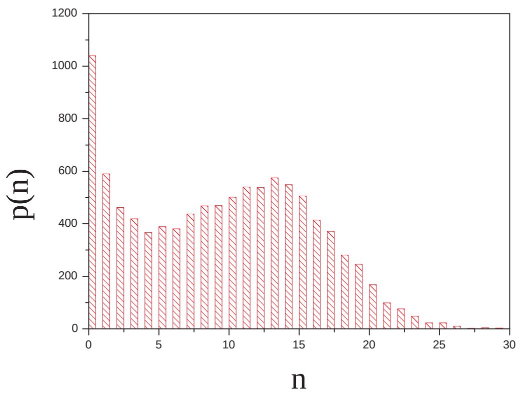

FIG. 4. Plot of $p(n)$ versus n obtained from stochastic simulation using Gillespie algorithm with the rate constants $k_1 = 1$, $k_2 = 6$, $k_a = 10$, $k_d = 1$, $J_p = 16$ and $k_p = 1$ (same as the long dashed curve of Fig. 3). For $k_p = 1$, $s_i = k_i\,(i = 1, 2, a, d)$ and $b_1 = J_p$.

The plot of $p(n)$ versus n for different values of $s_i\,(i = 1, 2, a, d)$ with $b_1 = 16$ is shown in Fig. 3. Different curves in Fig. 3 show that the distributions of protein levels are bimodal in different parameter regions with $s_i\,(i = 1, 2, a, d) \geqslant 1$. The binary response can also be observed in a region of parameter values with $s_i\,(i = 1, 2, a, d) < 1$ (not shown). The binary responses in the activator-repressor system cannot be observed for s_1, $s_d \geqslant 2$ (simultaneously). Figure 4 shows the binary response in protein levels obtained from stochastic simulation using Gillespie algorithm [23] for the biochemical reactions shown in Fig. 2 for the rate constants $k_1 = 1$, $k_2 = 6$, $k_a = 10$, $k_d = 1$, $J_p = 16$ and $k_p = 1$ (For $k_p = 1$, $s_i = k_i\,(i = 1, 2, a, d)$ and $b_1 = J_p$).
To understand the origin of bimodal distribution in protein levels in the present scenario we calculate the components of probability density function $p_i(n)(i = 1, 2, 3)$ in the steady state. Using Eqs. (9) and (11) one can easily obtain $p_3(n)$ and is given by

$$p_3(n) = C\,\frac{b_1^n\,\Gamma(g_1+n+1)\,\Gamma(g_2+n+1)\,\Gamma(h_1)\,\Gamma(h_2)}{n!\,\Gamma(h_1+n+1)\,\Gamma(h_2+n+1)\,\Gamma(g_1)\,\Gamma(g_2)}\,{}_pF_q(g_1{+}n{+}1;\ g_2{+}n{+}1;\ h_1{+}n{+}1;\ h_2{+}n{+}1;\ -b_1) \tag{13}$$

In the steady state, differentiating Eq. (8) n times w. r. t. z at $z = 0$ we have

$$p_2(n) = \frac{(s_d+b_1)}{s_a}p_3(n) - \frac{b_1^{n+1}\,\Gamma(g_1+n+2)\,\Gamma(g_2+n+2)\,\Gamma(h_1)\,\Gamma(h_2)}{k_p\,s_a\,n!\,\Gamma(h_1+n+2)\,\Gamma(h_2+n+2)\,\Gamma(g_1)\,\Gamma(g_2)}\times$$

$${}_pF_q(g_1+n+2;\ g_2+n+2;\ h_1+n+2;\ h_2+n+2;\ -b_1) \tag{14}$$

From Eq. (5) we have

$$p_1(n) = p(n) - p_2(n) - p_3(n) \tag{15}$$

where $p_2(n)$ and $p_3(n)$ are obtained from Eqs. (13) and (14) respectively.

Figure 5 shows the plot of total and component probability density functions $p(n)$ and $p_i(n)(i = 1, 2, 3)$ respectively versus n, the number of proteins, for the rate constants $s_1 = 1$, $s_2 = 6$, $s_a = 10$, $s_d = 1$ and $b_1 = 16$ (same as the dotted curve in Fig. 3). The bimodal distribution in protein levels is clearly the resultant of three unimodal functions $p_i(n)$ $(i = 1, 2, 3)$ (Fig. 5). This is also true for other bimodal curves in Fig. 3. From the rate constants used to obtain the bimodal

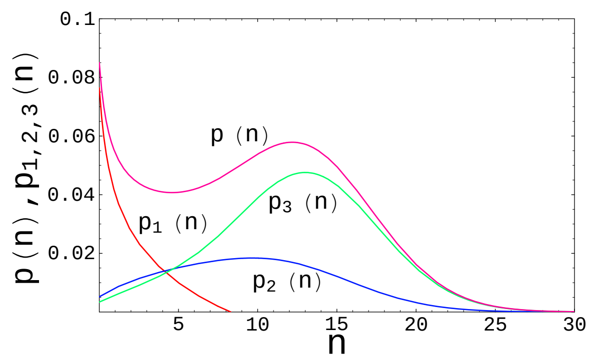

FIG. 5. Plot of $p(n)$ and $p_i(n)$ $(i = 1, 2, 3)$ versus n for the rate constants $s_1 = 1$, $s_2 = 6$, $s_a = 10$, $s_d = 1$ and $b_1 = 16$. The bimodal nature of the function $p(n)$ is the resultant effect of three unimodal functions $p_i(n)(i = 1, 2, 3)$.

distributions in Fig. 3, it is clear that the probability of occurrence of the gene in the G_1 and G_3 states are higher. Once it is in the G_1 state, the probability of transition to G_2 state is lower because of the lower value of s_1. On the other hand, at any instant of time, if the gene is in the G_2 state, there can be two possibilities: gene can switch either to the active state (G_3) or to the repressed state (G_1) depending on the amount of activators or repressors present in the system since activator molecules modulate the transition from G_2 to G_3 state and repressor molecules modulate the transition from G_2 to G_1 state. Higher values of s_a and s_2 make the G_3 and G_1 states more probable than G_2 and due to the lower values of s_1 and s_d, the gene spends most of the time either in the G_3 or G_1 state. Once the gene is in the G_3 state, the transition to G_2 state is rare because of the lower value of s_d. From the G_3 state of the gene, proteins are synthesized with rate constant J_p and there is enough time for the protein level to reach the steady value. This gives rise to high protein level in single cell and the peak in the distribution of protein level at higher value. This is clearly observed in the curve for $p_3(n)$ in Fig. 5. Now if the gene switches suddenly to G_2 state then protein level starts to decrease. The protein level keeps on decreasing as long as the gene is in the G_2 state or switches to the G_1 state. If the gene switches to the G_1 state the protein level decreases and reaches zero value. This gives rise to low/zero protein level in a single cell and the peak in the distribution of protein level occurs at low/zero value. This is observed in the curve for $p_1(n)$ in Fig. 5. From the G_2 state, the gene can also switch back to the G_3 state and this causes a rise of protein level again from an intermediate value. Therefore, there is a finite probability to observe the protein level at the intermediate value. The curve for $p_2(n)$ shows a finite value at the intermediate region of protein level (Fig. 5).

Rossi et al. examined whether an interplay of transcription factors can convert a graded to binary response in gene expression [12]. They designed an experiment in which the ratio of activator and repressor molecules that bind to the same promoter can be modulated by a single inducer molecule dox. Furthermore, the activator and repressor molecules bind the overlapping binding sites on the same promoter mutually exclusively. They analyze the graded and binary responses to the inducer molecule by flow cytometery in large population of individual cells. Three different cell populations viz. a dox regulated repressor ("repressor only"), a dox regulated activator ("activator only") and both ("activator+repressor") were generated in the experiment to study the role of positive and negative transcription factors. The flow cytometric analysis of the activator only and repressor only cell populations revealed a graded response (unimodal distribution) of GFP expression at all dox concentrations. The binary response (two distinct sub-populations) was observed in cells containing both activator and repressor molecules for a range of intermediate dox concentrations. With increasing dox level, the increase in the number of cells with maximal

level of GFP and decrease in the number of cells with low GFP level is observed. Therefore, an all-or-none (binary) response to the inducer level is observed in the experiment of Rossi et al. [12] when a combination of activator and repressor molecules act on the same promoter mutually exclusively. Moreover, since either factor independently produces a graded response, the binary response observed in cells with both the regulatory molecules is not due to a dominant effect of one factor over the other but rather to their combined effect.

Our theoretical analysis of the activator-repressor system does not explicitly include the activator and repressor numbers in the equations but are included in the rate constants s_a and s_2. The rate constant s_a increases with the increase in activator amount and s_2 increases with the increase in repressor amount. Now, let us assume that the numbers of both molecules can be controlled by a single inducer molecule like dox, as in the experiment of Rossi et al. [12], so that s_a increases and s_2 decreases with the increase of dox. Depending on the presence of regulatory molecules, the gene regulatory network can be divided into three categories: activator-only system (i.e., only activator molecules regulate the network), activator-repressor system (i.e., activators and repressors both regulate the network) and repressor-only system (i.e., only repressor molecules regulate the network). In presence of only activator molecules the three-state gene activation process reduces to the two-state one. Random switching then takes place only between G_2 and G_3 states. With the two-state gene activation process, the graded and binary responses are observed for $s_a, s_d > 1$ and $s_a, s_d < 1$ respectively [19]. In one hand, with $s_d > 1$ if s_a is varied from low to high value then unimodal responses are observed. On the other hand, with $s_d < 1$, if s_a is varied from low to high value then first unimodal (for $s_a < s_d$), then bimodal (for $s_a \simeq s_d$) and then again unimodal (for $s_a > s_d$) responses are observed [19]. Rossi et al. observed graded responses in activator-only system at all levels of inducer. To reproduce the experimental observations of Rossi et al. we choose the parameter region $s_a, s_d > 1$. Let us assume that initially there are only activator molecules (with low copy number) activating the gene transcription and s_d is fixed at 1.25. Now, with the gradual increase of inducer molecules dox in the system, s_a increases and the mean protein level also increases gradually. The probability distributions always remain graded (curves in the left column of Fig. 6) because the values of s_a and s_d satisfy the condition of unimodal/graded response ($s_a, s_d > 1$) for all values of dox [19]. Let us now consider the same regulatory network (same s_d) but with repressor molecules also present in the system. The gene can now switch between all three possible states and both the molecules compete for their binding site to take control of the gene transcription. Let us assume that initially there are large number of repressors i.e., s_2 is large and small number of activators i.e., s_a is low. With the gradual increase of dox molecules in the system, s_2 decreases and s_a increases gradually and simultaneously i.e., inhibition effect decreases and activation effect increases simultaneously. This causes the conversion of unimodal (for low dox i.e., low s_a and high s_2) to bimodal (for intermediate dox i.e., intermediate s_a and s_2) and then again unimodal (high dox i.e., high s_a and low s_2) distribution of protein levels (right column of Fig. 6). The gradual increase in the inducer level causes a discontinuous change in the mean protein level. Therefore, the response is bimodal/binary as the mean protein level is not a continuous function of inducer but has only low and high values. These results (Fig. 6) are in qualitative agreement with the experimental observations of Rossi et al. [12] for activator-only and activator-repressor systems.

Rossi et al. [12] also observed the graded response when only repressor molecules regulate the gene transcription. To reproduce the experimental observation for repressor-only case in the present scenario, one has to consider the basal rate of protein synthesis from the unregulated state of the gene (G_2). With the basal rate of protein synthesis, say J_0 ($J_0 < J_p$), from the G_2 state, the generation of graded response for repressor-only case is quite similar to that of the activator-only case discussed above. In the presence of only repressor molecules in the system, the

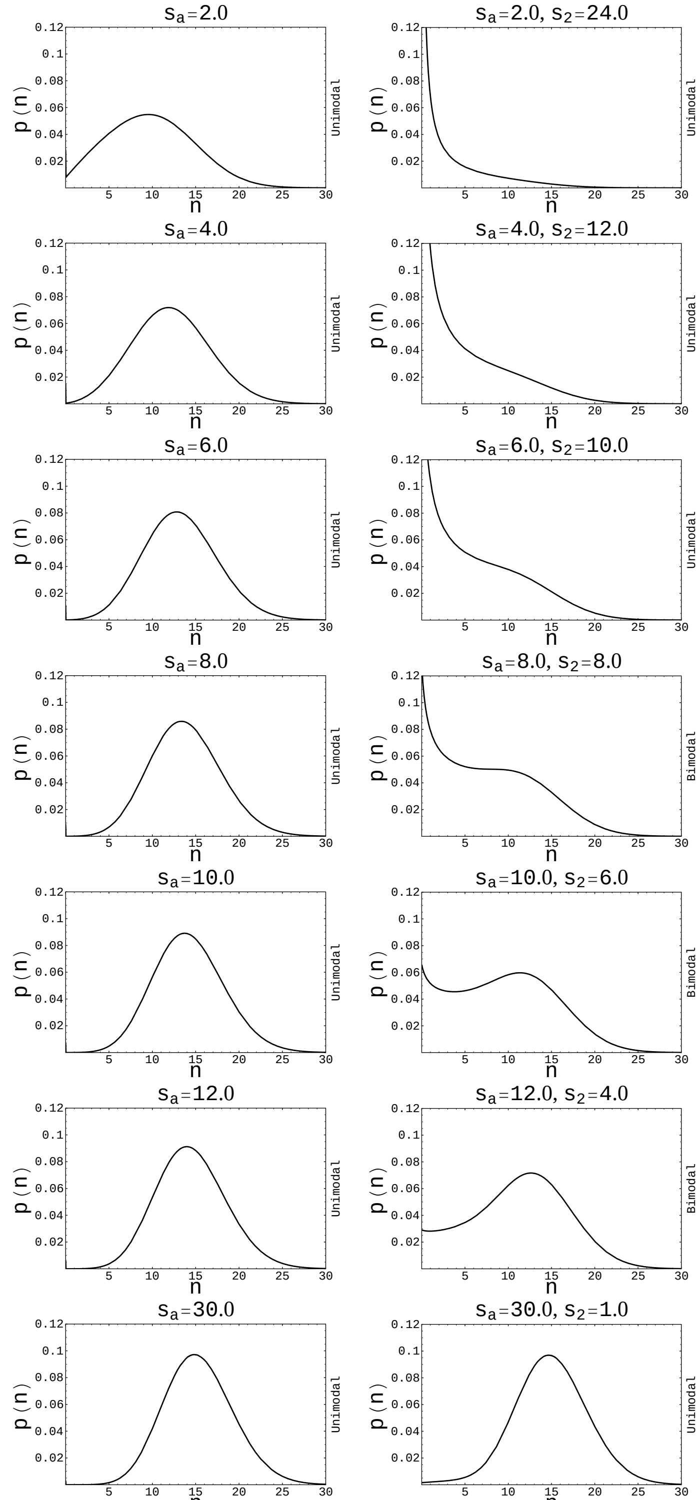

FIG. 6. Distribution of protein levels $p(n)$ versus n in activator-only system (left column) and activator-repressor system (right column) for different level of inducer molecules i.e., for different values of s_a and s_2. For the activator-only system s_2 is kept fixed at 1.25 and s_a is varied as mentioned on the top of the figures (figures in the left column). For the activator-repressor system s_1 and s_d are kept fixed at 1.25 and 1.25 (s_d is same as in the activator-only system) respectively and the different curves are drawn for different values of s_a and s_2 mentioned on the top of the figures (figures in the right column). For all curves the relative transcription rate constant $b_1 = 16$.

three-state gene activation process reduces to the two-state one i.e., the gene can switch randomly only between G_1 and G_2 states. Repressor molecules help in transition from G_2 to G_1 state. With the finite basal rate of protein synthesis from the G_2 state the response will be graded for $s_1 > 1$ and for all values of s_2 [19]. The initial value of the rate constant s_2 is large due to the presence of large number of repressor molecules in the repressor-only system, the response in this case will be unimodal since $s_1 = 1.25$ and s_2 is large. Now with the gradual increase of dox concentration, the rate constant s_2 decreases from a high to a low value but the response still remains graded due to s_1 being greater than one ($s_1 = 1.25$). Therefore, with the basal rate of protein synthesis from the unregulated state the graded response can also be observed for repressor-only system. But with the basal rate of protein synthesis from the state G_2 the derivation of exact analytical expression for the probability density function of protein levels for the activator-repressor system is very difficult. Though, with the help of stochastic simulation using Gillespie algorithm, it can be shown that the finite basal rate of protein synthesis from the state G_2 does not change our results qualitatively. The qualitative nature of the curves drawn in Figs. 3 and 6 will remain unchanged with the basal rate of gene expression from G_2 taken into account.

III. DISCUSSION

In this paper, we have studied an gene regulatory network where the positive and negative transcription factors regulate the gene transcription mutually exclusively. Both the molecules compete for their respective binding sites on the DNA to take control of the network (Fig. 1). The activator-repressor system is represented by a simple stochastic model where gene can be in three possible states viz. inactive/repressed, unregulated and active. An exact analytical expression for the probability density function of the protein levels in the steady state is derived and is a generalized hypergeometric function (GHF) (Eq. (12)). From the GHF, the bimodal distribution in protein levels is observed in a wide region of the parameter values. From the theoretical analysis, the experimental observation of Rossi et al. (i.e., the regulation only by activator molecules produce the graded response in the protein levels whereas binary responses are observed when both the activator and repressor molecules regulate the gene transcription by binding the promoter mutually exclusively) can be reproduced very easily. Here we have considered only the parameter region $s_i(i = 1, 2, a, d) \geq 1$ (Fig. 3). The binary response in protein level is more prominent for $s_i(i = 1, 2, a, d) < 1$ (not shown). This region is excluded from the present analysis because with $s_i(i = 1, 2, a, d) < 1$, the binary response can also be observed for two-state activator-only system [19]. But the experiment of Rossi et al. [12] observed only graded response when only activator molecules regulate the gene transcription. This experimental observation along with the theoretical prediction of graded response in gene expression [19] helps us to choose the parameter region for theoretical analysis. Rossi et al. observed the graded response also for repressor-only system. Here we have not considered the repressor-only case because this requires a basal rate of protein synthesis from the unregulated (G_2) state. The basal rate of protein synthesis from the unregulated state brings difficulties in the analytical tractability of the model. In the presence of the basal rate of protein synthesis from the unregulated state, it is very difficult to express the components of the generating functions $F_i(z)$ ($i = 1, 2, 3$) in terms of the total generating function $F(z)$ and therefore the Chemical Master Equations (CME) cannot be expressed by a single differential equation like Eq. (10). Again, the reduction of the CME into a single differential equation does not lead to the exact solution of CME because of the unavailability of the analytical solution of the higher order differential equation. This shows the limited scope and applicability of the generating function technique used here to solve the CME. The difficulty increases when the

regulatory networks consist of nonlinear feedback loops.

In the present analysis of activator-repressor system, we have combined the transcription and translation into a single step process. In the process of transcription mRNAs are produced from the active gene and then mRNAs are translated into proteins. Therefore, the steady state probability distribution (Eq. 12) derived here gives the correct description for mRNAs. The distribution in protein levels does not follow the bimodal mRNA distribution when the protein lifetime is longer than that of mRNA [24]. Despite the above limitations of the stochastic model, it contains important features necessary for an explanation of the binary response in an activator-repressor system and is graded in activator only system as observed in experiment [12]. The exact analytical result with three gene states is important and useful specially in the eukaryotic system. The gene activation of the complex eukaryotic system consists of many unknown number of rate limiting steps (chromatin remodeling, assembly of preinitiation complex etc.). The simplification of the complex gene activation process by the two-state one is the first approximation of the complicated biological process. The 'three-state' assumption may be considered as the second approximation of the stochastic gene activation-deactivation process.

The present analysis of the origin of binary responses in three-state model may be helpful to explain the bimodal distribution in transcriptional silencing [25]. In transcriptional silencing, Sir proteins (Sir 2-4) are the key structural components of silenced chromatin and under their regulation the silencer can be in two possible states: repressed and derepressed. The silencer helps to assemble the Sir protein complex. This process of assembling is not a single step process but rather consists of several reversible biochemical steps. The intermediate steps between the repressed and derepressed states of the chromatin make the effective rates of transitions very slow and these slow rate of transitions ultimately may lead to the bimodal distribution of protein levels from reporter gene.

Acknowledgement

This work is supported by the Minor Research Project Grant, UGC, India, under Sanction No. F. PSW-001/07-08 (ERO).

References

[1] A. Raj and A. van Oudenaarden, Cell **105**, 216 (2008).

[2] J. M. Raser and E. K. O'Shea, Science **309**, 2010 (2005).

[3] W. J. Blake, M. Kærn, C. R. Cantor and J. J. Collins, Nature **422,** 633 (2003).

[4] R. Karmakar and I. Bose, Phys. Biol. **3,** 200 (2006).

[5] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman and A. van Oudenaarden, Nat. Genet. **31,** 69 (2002).

[6] S. Ramsey, A. Ozinsky, A. Clark, K. D. Smith, P. de Atauri, V. Thorssons, D. Orrell and H. Bolouri, Philos. Trans. R. Soc. London B. Biol. Sci. **361,** 495 (2006).

[7] M Thattai and A. van Oudenaarden, Proc. Natl. Acad. Sci. **98,** 8614 (2001).

[8] P. S. Swain, M. B. Elowitz and E. D. Siggia, Proc. Natl. Acad. Sci. **99,** 12795 (2002).

[9] M. Kærn, T. C. Elston, W. J. Blake and J. J. Collins, Nat. Rev. Gen. **6**, 451 (2005).

[10] A. Becskei, B. Seraphin, and L. Serrano, EMBO J. **20,** 2528 (2001).

[11] C. Tan, P. Marguet and L. You, Nat. Chem. Biol. **5**, 842 (2009).

[12] F. M. V. Rossi, A. M. Kringstein, A. Spicher, O. M. Guichert and H. M. Blau, Mol. Cell. **6,** 723 (2000).

[13] S. R. Biggar and G. R. Crabtree, EMBO J. **20,** 3167 (2001).

[14] G. Zlokarnik, P. A, Negulescu, T. E. Knapp, L. Mere, N. Burres, L. Feng, M. Whitney, K. Roemer and R. Y. Tsien, Science **279,** 84 (1998).

[15] T. A. Carrier and J. D. Keasling, J. Theor. Biol. **201,** 25 (1999).

[16] R. Karmakar and I. Bose, Phys. Biol. **4,** 29 (2007).

[17] T. B. Kepler and T. C. Elston, Biophys. J. **81,** 3116 (2001).

[18] J. R. Pirone and T. C. Elston, J. Theor. Biol. **226,** 111 (2004).

[19] R. Karmakar and I. Bose, Phys. Biol. **1,** 1 (2004).

[20] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas and S. Tyagi, PLoS Biology **4,** 1707 (2006).

[21] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. **105,** 17256 (2008).

[22] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, Amsterdam, North Holland (1992).

[23] D. T. Gillespie, J. Chem. Phys. **81**, 2340 (1977).

[24] S. Iyer-Biswas, F. Hayot and C. Jayaprakash, Phys. Rev. E. **79**, 031911 (2009).

[25] E. Y. Xu, K. A. Zawadzki and J. R. Broach, Mol. Cell. **23,** 219 (2006).